My Friend's True Self: Children's Concept of Personal Identity

Michaela Jirout Košová[a], Robin Kopecký[a], Pavel Oulovský[a], Matěj Nekvinda[b], Jaroslav Flegr[a]

[a] Department of Philosophy and History of Science, Faculty of Science, Charles University, Prague, Czech Republic

[b] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic (until June 2018)

Contact: Michaela Jirout Košová; michaela.kosova@natur.cuni.cz; Department of Philosophy and History of Science, Faculty of Science, Charles University, Viničná 7, 128 00, Prague, Czech Republic

Abstract:

Our study explores the folk concept of personal identity in the developmental context. Two hundred and seventeen Czech children participated in an interview study based on a hypothetical scenario about a sudden change in their friend, someone they know, or some other unspecified person. The children were asked to judge to what extent particular changes (from 6 categories of traits) would change the identity core of their friend or some other person on a 7-point scale. We introduced both positive and negative versions of the changes. Our data suggest that children considered moral traits connected to interpersonal relationships crucial for preserving personal identity. Memory connected to personal experiences also scored highly. On the other hand, a change in physical appearance seemed to have the least important impact on personal identity. Negative changes turned out to have a significantly

greater impact than positive changes in all categories, except physical. Possible effects of scenario and the participants' age and sex were further explored. We discuss the possible causes of the effect of negative moral change and the role of social dimension in the development of the concept of personal identity.

1. Introduction

1.1. Current Knowledge

We, as social beings, live in a world full of selves: each person has various typical characteristics, both physical and psychological, which make them who they are - some of them more, others less so. We see each other as 'bundles' of traits and dispositions, which helps us orientate within the social environment and choose wisely when entering new relationships. However, people change throughout their lives and although some changes are seen as natural, others can threaten the integrity of the self. Thus it seems that the 'bundles' that we are made up of have certain essential cores. Many thinkers have devoted a substantial part of their work to developing theories about personal identity, trying to determine 'from the armchair' what is the most essential aspect of the self and what has to stay intact in order for a person's identity to remain unbroken.

Together with the researchers who inspired our current work, we are sceptical about most of these armchair-philosophy approaches. Since the concept of self is so very dependent on the fact that we are social beings, it is best to look for the answer in a natural environment: between people untouched by highly abstract philosophical debates, whose concepts naturally stem from their everyday life in society. Personal identity is a question that matters to everyone and we lean towards the view that it is defined by the way people view it in everyday social interactions and practices (Prinz & Nichols, 2016, p. 449). Thus, we see experimental philosophy as the right approach in the quest to find out more about these issues because it enables us to get in contact with the real-world phenomenon, even though we do not deny the value of purely theoretical inquiry.

This paper addresses current experimental-philosophy research on children's concept of personal identity and its relation to morality. The way we define morality in this context is specific and follows previous studies on the topic, since, as we already suggested, the results

of the existing studies point to the fact that interpersonal relationships are especially important in the context of revealing the folk concept of personal identity (Strohminger & Nichols, 2014, 2015; Prinz & Nichols, 2016). The focus of the paper is therefore on the interpersonal dimension of morality, by which we mean personal traits and dispositions that directly affect interpersonal relationships (e.g., cruelty, selfishness, criminal behaviour, callousness on the one side, and empathy, generosity, selflessness, conscientiousness, kindness on the other side). We can also define our sphere of interest as a kind of *moral essentialism* (Heiphetz, Strohminger & Young, 2017, p. 745).

Earlier studies exploring these issues were mostly questionnaire studies of adult respondents. Their results revealed a strong tendency in the respondents to ascribe importance to moral traits (esp. traits connected to the social realm) in the context of preservation of personal identity. In their online studies launched in 2009, Prinz and Nichols explored the folk concept of diachronic identity and managed to show that moral continuity is far more crucial for preserving personal identity over time than memory, agency, or narrative ability. They explained their findings by suggesting that morality is above all a social phenomenon and thus is of extreme importance to us. Moral values are sustained by the community in which we live, and our success as social beings depends on how this community views us (Prinz & Nichols, 2016, p. 463).

The above-mentioned findings about the role of the social dimension in the concept of personal identity inspired further studies that managed to support this view. In their five online questionnaire studies with adult respondents, Strohminger and Nichols (2014) found strong support for their hypothesis of the *essential moral self*. The intuition of most of their participants turned out to be that moral traits constitute the core of personal identity. Similar results further supported the hypothesis in another study focusing on the intuitions of relatives of patients with neurodegenerative disorders (Strohminger & Nichols, 2015). Symptoms of

disorders that affect the moral (interpersonal) dispositions of a patient were viewed by respondents as more identity-breaking than other common symptoms, such as amnesia.

Strohminger, Knobe and Newman (2017) refined the concept of the self by suggesting that apart from the concept of the *self* as such, folk people also have the concept of *true self*, which is a kind of subset of characteristics that are the most crucial and deep-down constituents of a person's self (p.552). Based on various psychological research findings, they defend the hypothesis that the *true self* is generally perceived as moral and inherently good (pp. 552-554), which could possibly be a result of the way we conceptualise essences or our need to maintain social bonds (pp. 556-557).

Similarly, Heiphetz, Strohminger and Young (2017) found that the role of social bonds seems to be crucial for the folk concept of personal identity. Participants in their studies judged belonging to a group as important for them and mediation analyses revealed the relationship between the perceived importance of community and the type of moral belief they judged to be more important for preserving personal identity. Changes to widely shared moral beliefs would result in a more radical identity change than changes in controversial moral beliefs, because widely shared beliefs are more closely connected to relationships in the community (p. 758).

The importance of interpersonal relationships in judgements about personal identity comes forward also in research conducted by Tobia (2016), which focused on the direction of hypothetical change. Tobia takes into consideration relational theories of personal identity, which emphasize the role of interpersonal relationships – an approach that is often overlooked in most philosophical theories (p. 38). In his study he broadened the classical scenario and found that respondents presented with the reversed Phineas Gage scenario (moral improvement) tended to claim that he is still the same person after the accident, whereas in the original scenario (moral deterioration) the opposite tendency was the case. Both changes are

equally radical, but the negative version leads more often to conclusions that the identity of the person in question was broken. Tobia suggests that the size of the change does not play the main role here and the reason why responses to opposing scenarios differ so greatly is connected to the fact that the self is viewed as essentially good (pp. 39-40). In his earlier paper (2015) Tobia supported his theory by further examples (Parfit's *nobleman* thought-experiment and also examples from pop-culture and literature). He concludes that normativity (which is connected to the social dimension) is to be viewed as a crucial aspect of the personal identity concept (p. 404).

Heiphetz, Strohminger, Gelman and Young (2018) also reported that their adult respondents judged the change from good moral beliefs to bad moral beliefs to have a significantly more serious impact on preserving personal identity than the change in the opposite direction (bad moral beliefs to good moral beliefs). The difference was mediated by the perceived influence of both types of changes on friendships. However, the results of the same study on children aged 8-10 years did not show these effects, possibly because of the sample size or study design (p. 216).

The authors also conducted another study comparing children and adults (pp. 211-214). The respondents were supposed to judge how much a person would change after taking a pill that changes either their widely shared moral beliefs, controversial moral beliefs, memories, or preferences. Children reported, similarly as adults, that the person would change more after a change in their widely shared moral beliefs took place.

The difference between adults and children appeared in the impact of memories and preferences on perceived identity change – in this context, children found memories and preferences more important than adults (p. 214).

The role of memory is, no doubt, crucial for the question of personal identity and has always been considered in this context (the classic example is John Locke, 1690) and has come forward also in findings of modern authors who emphasize the role of psychological in contrast to physical characteristics (Nichols & Bruno, 2010). However, studies exploring the difference between the importance of memory and moral traits for preserving personal identity show that in this context moral traits surpass memory (Prinz & Nichols, 2016; Strohminger & Nichols, 2014, 2015), although not consistently in all studies (Heiphetz, Strohminger & Young, 2017). It is thus desirable for future personal-identity studies to consider the role of memory and explore its importance in comparison to morality.

The question *why* interpersonal moral traits are so important for us certainly has to do with the fact that we are social beings and view ourselves and others in the context of social relationships (Prinz & Nichols, 2016, p. 463). However, our understanding of the fact why good moral traits appear in the very core of personal identity might be deepened by pointing to our natural tendency to essentialize (Strohminger, Knobe & Newman, 2017, pp. 556-557).

Psychological essentialism is a tendency to view entities as having an inner essence hidden behind superficial traits (De Freitas et al., 2018, p. 138). According to De Freitas and colleagues, the true self concept is a result of these natural essentialist tendencies, since the traits that are usually ascribed to the true self are viewed as "immutable, discrete and inherent" (De Freitas et al., 2018, p. 138). The answer to the question why social traits are more important than other kinds of traits might lie in the normative and teleological nature of essentialism. People seem to think about the essences of various entities in terms of their purpose, a *telos* (Rose & Nichols, 2019), and, at the same time, the traits that are believed to support personal identity are predominantly normatively good (De Freitas, Tobia, Newman & Knobe, 2017, p. 398). This logic that connects normativity and teleology could also be applied to human beings: if the purpose of the music band is to make good-quality music and

the purpose of scientific papers is to carry valuable scientific information, then the purpose of a person is to be morally good (De Freitas, Tobia, Newman & Knobe, 2017, p. 397). This theory seems to be in perfect accord with the battery of studies showing that good moral traits play superior role in preserving personal identity and explains why the folk view of personal identity seems so optimistic: a view that everyone is good deep inside (see De Freitas & Cikara, 2018).

Also the question of what exactly the studies described in this section and other similar studies reveal about the lay concept of personal identity is still open. Recent debate shows that we can interpret the results as a tendency by respondents to view a change in moral traits as leading to a significant change rather than annihilation of the person in question (Starmans & Bloom, 2018). On the other hand, it is possible to argue for a more radical interpretation, saying that with moral deterioration comes the end of the person in question, in the sense of numerical identity (De Freitas, Cikara, Grossman & Schlegel, 2018). There are studies, such as those of Tobia (2015), which manage to disambiguate between numerical identity and similarity (Dranseika, 2017) and still reveal the effect that we find in other, conceptually less precise studies.

We do not believe that it is necessary to stick to either of the interpretations in order to appreciate the aforementioned research. These studies show us that the concept of personal identity is closely connected to the social dimension and appreciation of moral goodness, regardless of the precise formulation of the folk concept. What is more, it might be misleading to expect philosophical rigour from the folk. On the contrary, we believe that we should be prepared to embrace a certain vagueness when trying to describe folk concepts.

1.2. Original Study

The study presented in this paper focuses on children's intuitions about personal identity, but it differs from the aforementioned study on children (Heiphetz, Strohminger, Gelman, & Young, 2018) by introducing more categories of change, a wider age-range of interviewed children and a larger sample size. Our aim was to explore the importance of the moral category in comparison to 5 other categories (physical, memory, character, perception, cognition) in the perceived preservation of personal identity in children aged 5-15 and thus explore the *essential moral self* hypothesis in the developmental context. By including both directions of change, we also focused on the *true self* concept.

Based on the current research findings in the field we decided to test the following hypotheses (as listed in the study preregistration)[1]: 1. The impact of hypothetical changes in various categories of traits will differ and the changes in moral traits (traits which figure strongly in interpersonal relationships) will be rated as having the highest impact on the perceived change in personal identity. 2. When the hypothetical change is negative, the overall change in personal identity will be rated as more serious than in cases when the change is positive. 3. In the case of negative moral change, the overall change in personal identity will be rated as much more serious that in any other category of negative change. 4. The exploratory part of the study: the effect of age and sex in rating the impact of different categories of change on preserving personal identity.

2. Method

2.1. Participants

In June 2017 we conducted an interview study on Czech children and teenagers. The study was preregistered and approved by the IRB.[2] In our preregistered plan we set a stopping rule at 300 respondents in the age range 5-15 years old. The number was supposed to approximately match the number of participants in the online studies on adult respondents by Strohminger and Nichols (2014), which ranged from 79 to 318.

Altogether we interviewed 267 respondents. As reported in the preregistration, participants who did not show a sufficient understanding of the study question and who exhibited random answers even before the experimenter properly finished particular questions were filtered out already during the data collection. We excluded 28 participants based on this rule. One participant was excluded due to the parent's report of autism. We also interviewed 21 participants who were above age 15, since we agreed to go through the whole interview with all potential participants who showed an interest even after they had reported their age at the beginning.[3] These participants were excluded before the analysis. One participant aged 5 was excluded because we were not able to get more participants of this age. We thus set the age range to 6-15. We ended up with a sample of 217 children for the final analysis (56.4% female; age range 6-15; average age=11).

2.2. Procedure

Respondents were randomly recruited at a public family event popularizing science. After gaining informed consent from a parent and asking the parent to step aside, an interviewer informed the child participant that she was exploring the thinking of children and young people and was interested in what the participant thought about the following problem. She assured the child that none of the answers were right or wrong and the point was to find out what the child's own personal opinion was.

Once the child had agreed to this model, the interviewer introduced each participant to a scenario in which a person undergoes various changes after being closed in a special sci-fi chamber. A random group of participants (N=90) were asked about their *friend*, another group of participants (N=91) were asked about *some person in general*, and the rest of the participants (N=36) were asked less generally ("your friend, peer or someone else"- *someone they know*). Changes in all scenarios encompassed 6 categories: physical (appearance), cognitive (intelligence), moral (love for others and treatment of others), in character

(laziness), in memory (remembering life experiences) and in perception (vision). Both negative and positive versions of the changes were included in all categories. Altogether we formed fourteen questions which were mixed and presented in the following order: the person 1. becomes uglier, 2. gains a super-memory, 3. stops loving their friends, 4. becomes more industrious, 5. becomes blind, 6. becomes more beautiful, 7. forgets their life experiences, 8. becomes nicer to others, 9. becomes more stupid, 10. accepts as a friend someone they didn't like before, 11. becomes lazier, 12. gains much better eyesight, 13. becomes cruel to others, 14. becomes smarter. The respondents were asked to judge how much each of the changes would affect the person's identity core ("the most crucial aspect of the person which makes them who they really are deep inside") on a 7-point scale (0 – "they are still the same person and their most crucial aspect remains intact"; 6 – "they are not the same person anymore and have lost their most crucial aspect"). We indicated the scale by circles growing in size (the larger the circle, the larger the overall change - see supplementary material for details and exact formulations). We decided to use this method because we needed the scale to be comprehensible to children who have little or no experience with reading (see e.g. Okawa, 2008, p. 188). The relationship between the size of the circle and the size of the change was pointed out to the respondents at the beginning.

Once the participant had finished answering the questions, the interviewer thanked them for their cooperation and gave each respondent a badge as a reward.

The choice of within-subject model (with an exception of the between-subject vignette question) was drawn by the fact that we were only able to gain a restricted number of respondents varying across age categories. The fixed order of questions also reflects the restricted conditions of the study – a face-to-face interview without the use of computer technology.

2.3. Differences between Preregistered and Realized Protocol

After the preregistration we divided the vignette into three scenarios: *friend*, *someone you know*, and *some person in general,* in order to explore the role of closeness of interpersonal relationships.

We adjusted the planned age range (5-15) to 6-15 years old, since we had only managed to recruit one 5-year-old.

To test hypothesis 3 we used one-sided paired t-tests (just as in the case of hypothesis 1 and 2). Due to a mistake we failed to mention this test in the preregistered protocol.

In the exploratory part of our analysis plan we added one more test – we used Welch's two-sample two-sided t-tests (Holm correction applied) in order to explore the differences between the relative importance of the moral category in comparison to the other categories in-between age categories, scenario categories and sexes.

3. Results

Prior to all data analyses, we calculated the z-score[4] for each participant to control for between-subject differences in rating and to transfer the ordinal scale to a continuous scale. We set the alpha level to 0.05 and applied Holm correction for all multiple tests.

We performed five one-sided paired t-tests to test hypothesis 1 concerning differences between the perceived importance of change in the moral category and changes in each of the five other categories (physical, cognitive, memory, character, perception). Data analyses showed that respondents considered a change in the moral traits (M = 0.5, SD = 0.35) to have a significantly greater impact on personal identity than a change in any other category of traits, esp. physical (M = -0.65, SD = 0.57, t(216) = 21.91, p < 0.0001), then perception (M = -0.38, SD = 0.58, t(216) = 16.2, p < 0.0001), character (M = -0.12, SD = 0.46, t(216) = 15.09, p <0.0001), cognition (M = -0.04, SD = 0.45, t(216) = 13.15, p < 0.0001) and finally memory

(M = 0.19, SD = 0.59, t(216) = 6.24, p < 0.0001) (see figure 1, and also fig. 1 in the Appendix).

[FIGURE 1 here]

We further performed six one-sided paired t-tests to test hypothesis 2 concerning differences between the perceived importance of negative and positive versions of a change in each of the 6 categories. The results showed that a negative change was perceived as having a significantly greater impact on the preservation of personal identity than a positive change in all categories, except physical. The most salient difference showed itself in the rating of the relevance of the change in the moral category (positive: M = -0.07, SD = 0.56, negative: M = 1.08, SD = 0.55, t(216) = 19.66, p < 0.0001). Also super-memory gain (M = -0.28, SD = 0.85) and the corresponding memory loss (M = 0.67, SD = 0.89) were rated very differently, with memory loss scoring significantly higher (t(216) = 10.91, p < 0.0001). On the other hand, when it comes to a change in physical appearance, negative and positive versions (becomes uglier: M = -0.66, SD = 0.78 vs. becomes more beautiful: M = -0.65, SD = 0.75) received an almost equal rating and thus the result was not anywhere near significant (t(216) = -0.19, p = 0.57) (see figure 2, and also fig. 2 in the Appendix).

[FIGURE 2 here]

To test hypothesis 3 we analysed a subset of data on negative versions of changes in each category and performed five one-sided paired t-tests to test the difference between the perceived importance of a negative moral change and a negative change in all the other categories. Analysis showed that a negative moral change (M = 1.08, SD = 0.55) was rated as significantly more important than a negative change in any other category (in all cases p < 0.0001).

As a part of additional exploratory research, we also tested the difference between age groups of 6-8, 9-11 and 12-15 years (according to standard growth stages, see e.g. age periods in Gibbs, 2014, p. 74), sexes, and scenarios. As planned in preregistration, we performed the ANCOVA test with *change in personal identity* as the output variable and the direction, category, sex, age and the direction-category interaction as predictors. The effect of the category ($F(5, 3022) = 156.26$, $p < 0.0001$, $\eta_p^2 = 0.205$), direction ($F(1, 3022) = 344.57$, $p < 0.0001$, $\eta_p^2 = 0.102$) and category-direction interaction ($F(5, 3022) = 42.61$, $p < 0.0001$, $\eta_p^2 = 0.066$) proved to be significant.

In the resulting graph of the age categories we observed that there is a change in the relative importance of the moral category in comparison to the other categories of change across age, that is, relative importance of the moral category grows with age. Thus, we concluded that comparing the values of scores of the same categories between age groups would not be the right way to approach the data. Instead, we decided to focus on comparing differences in the scores of the moral and non-moral categories (i.e., the mean score of the moral category minus the mean of the scores of all the other categories). Then we applied Welch's two-sample two-sided t-tests (Holm correction applied) to test the statistical significance of the difference between the relative importance of the moral category in comparison to the other categories in different age groups, which proved to be significant between the 1st age category and the 2nd age category ($p = 0.002$) and the 1st age category and the 3rd age category ($p < 0.001$). It thus seems that the relative importance of the moral category grows with age (see figure 3). The most prominent change seems to take place around the age of nine, and it is especially obvious with regard to the contrast between the moral and the physical category, while memory remains close to the moral category (see fig. 3-7 in the Appendix).

[FIGURE 3 here]

We also applied the same test in order to explore differences between sexes. The relative importance of the moral category was higher in the case of female respondents ($p = 0.004$). Girls considered the moral category as having a greater impact on the preservation of personal identity in comparison to the other categories than boys (see figure 4). The most prominent difference shows itself in the contrast between the importance of the moral vs. physical category, while memory stays very close to the moral category (see fig. 8-12 in the Appendix).

[FIGURE 4 here]

We also tested this effect in the case of scenarios. Again, Welch's two-sample two-sided t-tests (Holm correction applied) revealed that there is a statistically significant difference in the relative importance of the moral category in comparison to the other categories between the *friend* scenario and *some person in general* scenario ($p < 0.001$). The more personal the scenario, the greater the relative importance of the moral category (see figure 5). This effect seems to be much more prominent in the two older age categories, and again seems to be obvious especially in the contrast between the moral and the physical category, while memory remains close to the moral category (see fig. 13-15 in the Appendix).

[FIGURE 5 here]

4. Discussion

Our results support the view that moral continuity is crucial for the preservation of personal identity and that the *essential moral self* and the *true self* hypotheses go in the right direction. The preference of moral traits in the concept of the self was already present in children and teenagers. Moral traits (or more precisely, traits that have an important role in the social context) proved again to be central to the folk concept of personal identity.

Personal identity was perceived to be broken especially when the negative moral change was the case. This could be explained by the fact that the questionnaire was based on respondents' intuitive assessment. This means that the answers are not strictly rational (i.e. not a result of moral reasoning accompanied by deeper reflexion), but also – and mainly – based on emotional (quick, automatic and intuitive) evaluation. It is consistent with the social intuitionist model (Haidt, 2001). The relatively greater importance of the negative change (in comparison to the positive change) means that the weight of the negative feeling in case of a negative change is more significant than the weight of the positive feeling related to a positive change. This phenomenon may be potentially explained by a feeling of danger, as a negative change of moral traits is associated with asocial and selfish behaviour and could thus lead to social conflict.

This leads us to a further suggestion that the phenomenon of the negative change's greater importance may also have a social meaning, which would support our view that the folk concept of the self is socially determined. Moral traits are the traits that hold society together. A person evaluated as *bad* displays a less pronounced tendency to integrate into society. Their interests may even go against the interests of society and be destructive to it. Behaviour motivated by such interests is generally perceived negatively by the society in which the children live. This indicates that the folk concept of the self operates with an idea of the self that is substantially pro-socially oriented and that this orientation may be observed even in the case of children. This theory is also in accordance with evolutionary theories of morality (Haidt, 2007; Keefer, 2013) and the idea that our concepts stem naturally from our need to be able to orientate within social interactions and maintain good social bonds.

The *social* explanation could be deepened by considering the points made by psychological essentialism. Natural tendency to essentialize entities based on their hidden traits might have the above described benefits of maintaining good social bonds, especially

when the entity in question (in this case a person or a group of people) is viewed as good deep inside. The positive effect of the true self concept on the intergroup relationships was also demonstrated experimentally (De Freitas & Cikara, 2018). This might be related to the normative and teleological nature of essentialism. If people indeed see essences in the context of purposes (Rose & Nichols, 2019) and these purposes are normatively good (De Freitas, Tobia, Newman & Knobe, 2017, p. 397), then it makes perfect sense that a human being should be viewed as essentially moral and well disposed towards others, and thus capable of forming good quality interpersonal relationships. Well-functioning society is, after all, naturally desirable for any social species. A person is seen as losing his or her identity in the case of negative change because they have diverted from their true self. Not being recognized as the same person might even be perceived as a kind of social punishment.

The tendency of the respondents to judge moral change as causing a more radical disruption in personal identity when their friend is the person undergoing the change further accentuates the proposed view. Friends are persons to whom we are emotionally related and who have a high position in the hierarchy of human relationships that reflects patterns in human altruistic behaviour (Ma, 2013, p. 3). A change in a friend's moral attitudes potentially leads to the loss of this mutually valued relationship, which is accompanied by unpleasant feelings. We propose that these feelings cause relatively greater sensitivity to a change in a friend's moral traits and thus a more radical judgement concerning a rupture in personal identity when it comes to morality. Our results support the theory that emphasizes the substantiality of moral traits - the *essential moral self*, together with theories that do justice to the social aspect of personal identity.

Data also reflect the relative importance of autobiographical memory. Our findings are in agreement with the previous conclusions reached by researchers within this field (Wilson & Ross, 2003; Bluck et al., 2010, Demiray & Janssen, 2015). The difference, however, lies in

our concluding that the importance of moral traits is relatively greater than that of autobiographical memory, which is in accord with findings from the earlier studies already mentioned in section 1.1. of this paper. We believe that this is not in contradiction with the more widespread theory that emphasizes the significance of autobiographical memory for the establishment of personal identity. However, based on both previous research and our own research we may conclude that moral traits play an even more essential role here. It could also be the case that memory of life experiences is closely connected to personal memories of interpersonal relationships, which again supports theories which put forward the social aspect of the folk concept of personal identity.

In deep contrast to moral traits, physical appearance played the least important role in the children's concept of personal identity, which is again highly in favour of the claims that psychological traits surpass physical traits when it comes to personal identity intuitions.

The observed effect of age can be a result of a specific aspect of moral development as it is described in the theory of relationship between *moral self* and *moral identity* by Kingsford and colleagues (Kingsford, Hawes, & de Rosnay, 2018). In early childhood, children gain a concept of themselves as moral persons (a *moral self*) based on repeated experiences of their conduct in morally coloured social situations (p. 656). However, this concept or self-knowledge lacks certain generalising and evaluative dimensions: the children are able to recognize, which types of behaviour are consistent with their own, but the step towards general awareness of the kind of moral person they are doesn't take place yet (p. 657). This type of self-knowledge thus differs from the *moral identity* of adolescents and adults, where this higher-order reflexion leads to a more consistent and evaluative moral self-image (p. 658). The crucial difference shows itself in attempts to define what makes an action *moral.* One of the answers points to a *second-order desire* to pay justice to the kind of moral person one is that appears from the age of eight, which is in contrast to mere acknowledging

and respecting moral norms that is characteristic of younger children (five to seven years old) (pp. 659-660). To reach the ability to be "genuinely and independently self-evaluative" requires cognitive capabilities that don't usually appear before the ascent of middle childhood (eight to twelve years). Children in their middle childhood even start to be able to compare their specific personal traits to others, especially their peers (p. 661). It is thus possible that older children understand moral traits as much more integrative part of personal identity and understand the concept of the coherent moral self much better than younger children. This explains the specific effect between the youngest age category (6-8) and the middle age category (9-12).

As social beings, children are constantly under the influence of their parents and teachers, who instil in them the moral values of their community and present them with their idea of correct social conduct. Children are repeatedly told that it is not right to judge people by the way they look, that it is important to treat other people right, not hurt them and be polite and nice. Even though authorities play a crucial role in the development of the moral self, children are active participants of this process and their responses to the parental influence are domain specific (Smetana & Jambon, 2017, pp. 133-134). Our results suggest that despite the fact that children are also being reprimanded for their laziness, this trait does not seem to have a significant impact on personal identity in their view. Therefore, some natural appreciation of moral traits seems to play an important role, perhaps due to already mentioned teleological essentialist tendencies.

The observed effect of sex may be related to the age effect, since girls are usually developmentally ahead of boys. It may also be explained via the tendency of girls to ascribe greater importance to social relationships than boys, which is partially encouraged by upbringing based on certain cultural stereotypes about sexes (Eagly & Wood, 2017).

In conclusion, children judged the true self of a person to be well disposed towards other people and inherently good. A change in the direction towards anti-social behaviour or in loss of memory bonds to others was considered as much more identity-breaking than a change in any of the other categories. Our study with respondents from a central European country further supports the view that the *true self* concept is cross-culturally stable on a more abstract level (good moral traits are central to it), even though evaluation of particular acts as good or bad may vary across cultures to some extent (Strohminger, Knobe, & Newman, 2017, p. 554).

We are aware that our research has certain limitations. Even though we mixed both the categories and the positive and negative versions of the questions so that children couldn't easily register our focus on the moral items, the order of the items was fixed. This could have an effect on the answers. Also the range of traits used in our vignette is considerably narrow, since we needed the interview to be brief due to children's short attention span. The children's responses to the interviewer – a stranger and an adult figure – could also play a role in their tendency to answer in a certain way (to meet the adult's expectations, perhaps). However, we used simple language in order to make sure that the children understood the questions and observed their feedback. In several cases, the children even spontaneously revealed the reasoning behind their answers and thus showed to what extent the question resonates with their intuitions (for more details, see data).

In future research it would be apposite to widen the range of categories and the number of items in each category. It would also be convenient to randomize the order of the questions properly and use between-subject model to control for the possible order-effects. A much larger study (more participants, more interviewers and a use of computer randomiser) would be needed to ensure that the effects we reported are indeed present and apply more generally. Confirmation of our findings concerning the age and scenario effects would require

testing our exploratory questions again, while formulating hypotheses in advance and running confirmatory research. However, we believe that each study is only revealing the phenomenon from a certain specific point of view and we need a battery of similar and at the same time slightly varied studies to hint at a wider picture. Being able to turn to the large number of existing results, we nevertheless dare to assert that the proposed existence of the concepts such as *the essential moral self* and *the true self* has already gained very robust support.

Acknowledgment:

MJK and RK developed the study concept and were responsible for the study design and data collection. MJK and JF were responsible for the study protocol and its preregistration. MN performed the data analyses. MJK, RK, PO and MN interpreted the results. MJK and PO drafted and all the authors revised the manuscript. All the authors approved the final version of the manuscript for submission.

Notes:

1. "Lay concept of personal identity and the importance of moral and interpersonal traits" preregistered on 8th June 2017. Link: https://osf.io/qj7k6/

2. IRB Charles University, Faculty of Science, approval number: 2017/14

3. Parallel research with another interviewer on the topic of trolley problems took place nearby at the family event. Respondents (age-range 6 - 18 years) were encouraged, but not obliged to participate in both studies.

4. Alternative version of the analyses without computing the z-scores is available here: https://osf.io/qfvu3/

Graphs showing the original scores without any transformations are available in the Appendix.

Funding Details:

Disclosure statement:

Notes on Contributors:

Michaela Jirout Košová is a PhD candidate at the Faculty of Science of Charles University. Her domain of interest is philosophy of mind and throughout her studies she has published two papers in Czech philosophical journals. She also conducted experimental-philosophy research as a team member of grant projects at Charles University. *The Philosophical Journal* – a renowned philosophical journal in the Czech Republic - recently accepted her paper introducing personal identity research and experimental philosophy. She is also a member of *The Karel Čapek Center for Values in Science and Technology*.

Robin Kopecký is a PhD candidate at the Faculty of Science of Charles University. His domain of interest is experimental philosophy and evolutionary biology. He is a member of the laboratory of evolutionary biology, participates in a number of grant projects and conducts research in evolutionary psychology and experimental philosophy. His paper on ethics and autonomous vehicles was accepted by *The Philosophical Journal* in the Czech Republic. He

is a coordinator and member of *The Karel Čapek Center for Values in Science and Technology*.

Pavel Oulovský is a PhD candidate at the Faculty of Science of Charles University. He focuses on depth psychology and existential psychology. He is a member of the Czech Daseinsanalytical Society and Austrian Daseinsanalytical Institute. He was a team member of a grant project at the Charles University, where he analysed psychological diagnostic tools for interpersonal behaviour. His paper (co-authored by Jirout Košová and Kopecký) on psychometric properties of the Circumplex Scales of Interpersonal Values in the Czech population has recently been accepted for publication by a renowned Czech psychology journal *Czechoslovak Psychology*.

Matěj Nekvinda earned his MA at the Faculty of Mathematics and Physics of Charles University in Prague. His main domain of interest is statistics in natural sciences.

Jaroslav Flegr is a tenured professor of biology at the Faculty of Science of Charles University and a researcher at the National Institute of Mental Health in the Czech Republic. He specialises in parasitology, evolutionary biology and biological psychiatry. In addition to some 150 scholarly papers, he has also published four monographs.

ORCID: Michaela Jirout Košová: https://orcid.org/0000-0002-0441-8426

Data Availability: https://osf.io/cjuf9

References:

Bluck, S., Alea, N., & Demiray, B. (2010). You get what you need: The psychosocial functions of remembering. In J.H. Mace (Ed.), *The act of remembering: Toward an understanding of how we recall the past* (pp. 284–307). Malden, MA: Wiley-Blackwell.

De Freitas, J., Tobia, K., Newman, J.E., & Knobe, J. (2017). Normative Judgements and Individual Essence. *Cognitive Science* 41, pp. 382-402.

De Freitas, J., & Cikara, M. (2018). Deep down my enemy is good: Thinking about the tue self reduces intergroup bias. *Journal of Experimental Social Psychology*, 74, pp. 307-316.

De Freitas, J., Cikara, M., Grossmann, I., Schlegel, R. (2018). Moral Goodness Is the Essence of Personal Identity. *Trends in Cognitive Sciences*, Letter, 22(9), pp. 739-740.

De Freitas, J., Sarkissian, H., Newman, G. E., Grossman, I., De Brigard, F., Luco, A., Knobe, J. (2018). Consistent Belief in a Good True Self in Misanthropes and Three Interdependent Cultures. *Cognitive Science*, 42, pp. 134-160.

Demiray, B., & Janssen, S. M. J. (2015). The Self-enhancement Function of Autobiographical Memory. *Applied Cognitive Psychogy*, 29, pp. 49-60.

Dranseika, V. (2017). On the Ambiguity of 'the Same Person'. *AJOB Neuroscience*, 8(3), pp. 184-186.

Eagly, A., & Wood, W. (2017). Gender Identity: Nature and Nurture Working Together. *Evolutionary Studies in Imaginative Culture*, 1(1), pp. 59-62.

Gibbs, J. C. (2014). Moral development and reality: Beyond the theories of Kohlberg, Hoffman, and Haidt (3rd ed.). Oxford University Press.

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, 108(4), pp. 814-834.

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science,* 316(5827), pp. 998-1002.

Heiphetz, L., Strohminger, N., & Young, L. L. (2017). The Role of Moral Beliefs, Memories, and Preferences in Representations of Reality. *Cognitive Science,* 41, pp. 744-767.

Heiphetz, L., Strohminger, N., Gelman, S., & Young, L. (2018). Who am I? The role of moral beliefs in children's and adults' understanding of identity. *Journal of Experimental Social Psychology*, 78, pp. 210-219.

Keefer, W. M. (2013). Understanding morality from an evolutionary perspective: challenges and opportunities. *Educational Theory*, 63(2), pp. 113-131.

Kingsford, J. M., Hawes, D. J., & de Rosnay, M. (2018). The moral self and moral identity: Developmental questions and conceptual challenges. *British Journal of Developmental Psychology*, 36, pp. 652-666.

Locke, J. (2009). *Essay Concerning Human Understanding*. WLC books. (Original work published 1690)

Ma, H. K. (2013). The moral development of the child: an integrated model. *Frontiers in Public Health*, 1:57.

Nichols, S. & Bruno, M. (2010). Intuitions about personal identity: An empirical study. *Philosophical Psychology*, 23(3), pp. 293-312.

Okawa, J. B. (2008). *Considerations for the Cross-Cultural Evaluation of Refugees and Asylum Seekers*. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of Multicultural Assessment: Clinical, Psychological, and Educational Applications (Third edition)*. San Francisco, CA: John Wiley & Sons, Inc.

Prinz, J. & Nichols, S. (2016). *Diachronic identity and the moral self*. In J. Kiverstein (Ed.), *The Routledge Handbook of Philosophy of the Social Mind* (pp. 449-463). Abingdon: Routledge.

Rose, D. and Nichols, S. (2019), Teleological Essentialism. Cognitive Science, 43: e12725. doi:10.1111/cogs.12725

Smetana, J. G. & Jambon, M. (2017). *Parenting, Morality and Social Development*. In Ch. C. Helwig (Ed.), *New Perspectives on Moral Development* (pp. 121-139). Routledge.

Starmans, C. & Bloom, P. (2018). Nothing personal: what psychologists get wrong about identity. *Trends in Cognitive Sciences*, 22, pp. 566-568. DOI:https://doi.org/10.1016/j.tics.2018.04.002

Strohminger, N. & Nichols, S. (2014). The essential moral self. *Cognition*, 131, pp. 159-171.

Strohminger, N. & Nichols, S. (2015). Neurodegeneration and Identity. *Psychological Science,* 26(9), pp. 1469 - 1479.

Strohminger, N., Knobe, J., & Newman, G. (2017): The True Self: A Psychological Concept Distinct From the Self. *Perspectives on psychological science,* 12(4), pp. 551-560. doi: 10.1177/1745691616689495.

Tobia, K.P. (2015). Personal identity and the Phineas Gage effect. *Analysis,* 3(75), pp. 396-405. doi:10.1093/analys/anv041

Tobia, K.P. (2016). Personal Identity, Direction of Change, and Neuroethics. *Neuroethics* 9(37), pp. 37-43. https://doi.org/10.1007/s12152-016-9248-9

Wilson, A. & Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. *Memory*, 11(2), pp. 137–149.
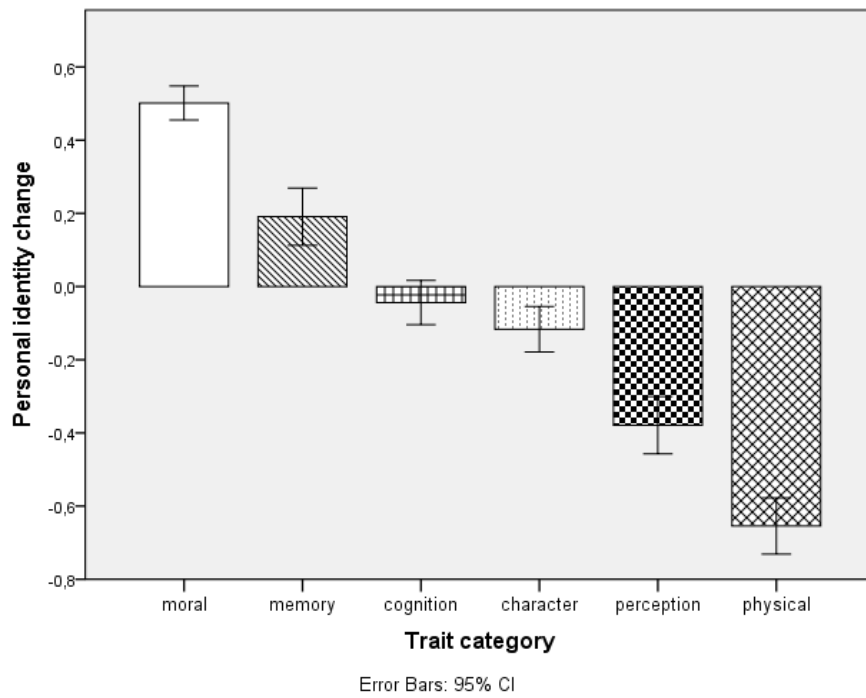
Fig. 1
Comparison of the perceived change in personal identity in different categories of traits.
Respondents considered a change in the moral traits to have a significantly greater impact on personal identity than a change in any other category.
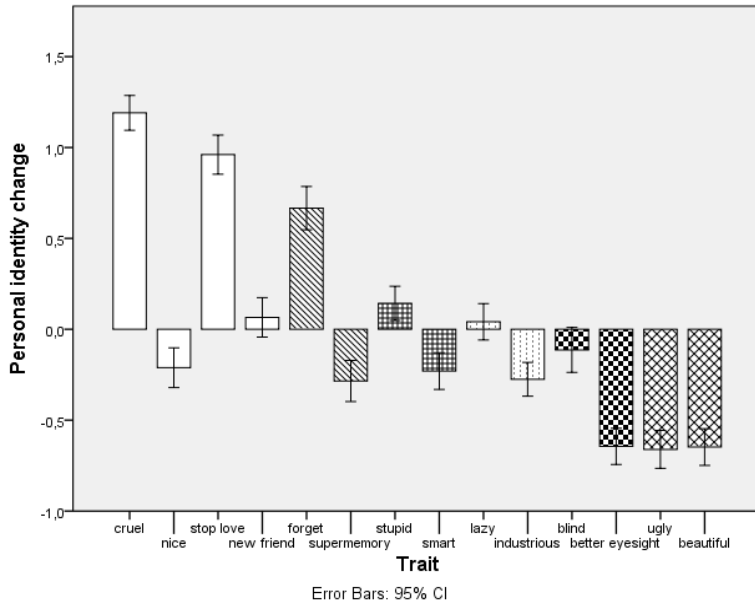
Fig. 2
Differences between the perceived importnace of negative and positive versions of a change.
The negative change was perceived as having a significantly greater impact on the preservation of personal identity than a positive change in all categories, except physical.
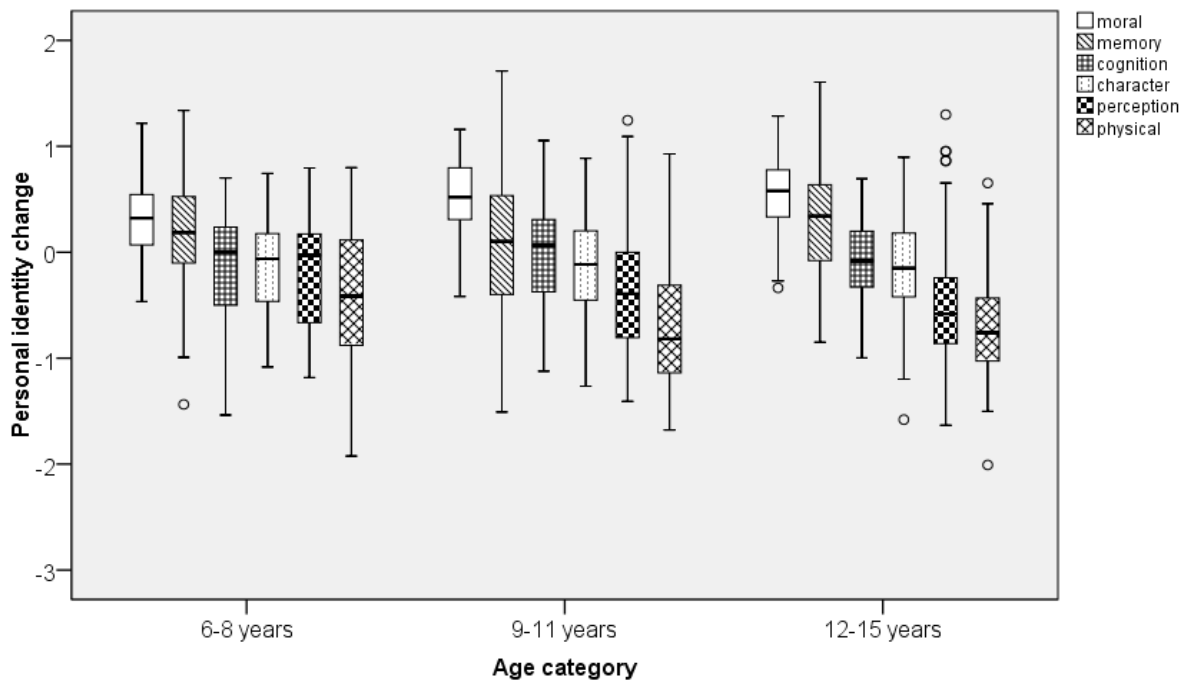
Fig. 3
The effect of age on the perceived change in personal identity.
The relative importance of the moral category grows with age.
The bottom end of the whisker is minimum value. The top end of the whisker is maximum value.
The bottom of the box is 25th percentile. The top of the box is 75th percentile.
The horizontal line inside the box is median. The dots represent outliers.
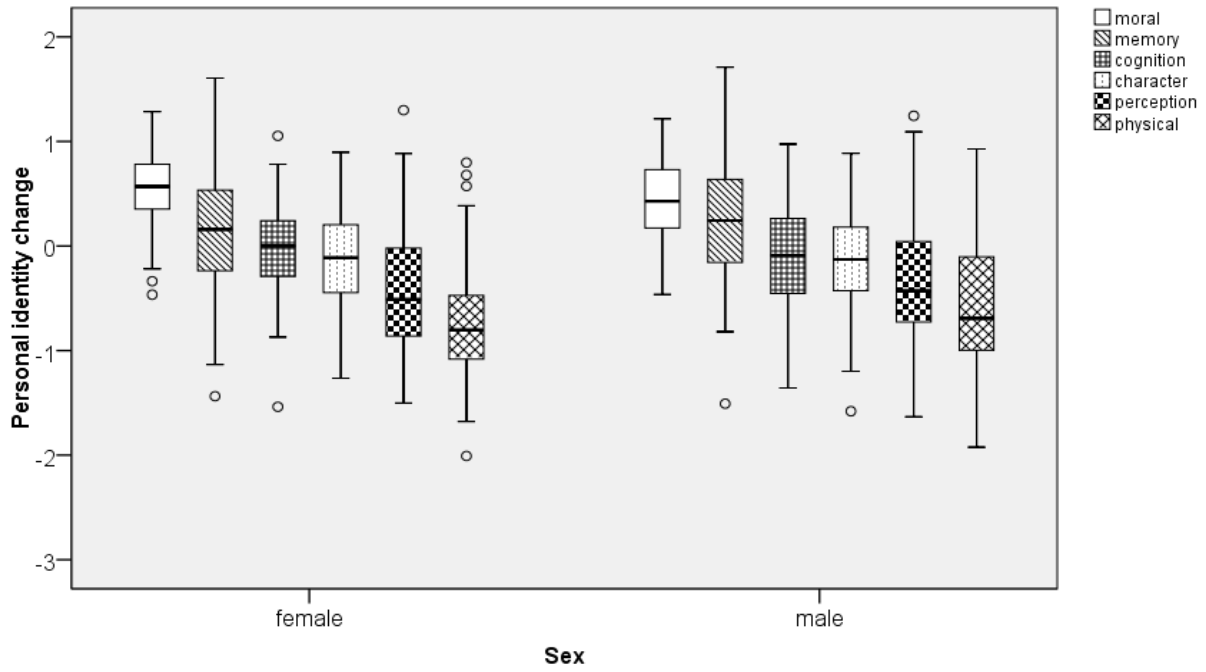
Fig. 4
The effect of sex on the perceived change in personal identity.
Girls considered the moral category as having a greater impact on the preservation of personal identity in comparison to the other categories than boys .
The bottom end of the whisker is minimum value. The top end of the whisker is maximum value.
The bottom of the box is 25th percentile. The top of the box is 75th percentile.
The horizontal line inside the box is median. The dots represent outliers.
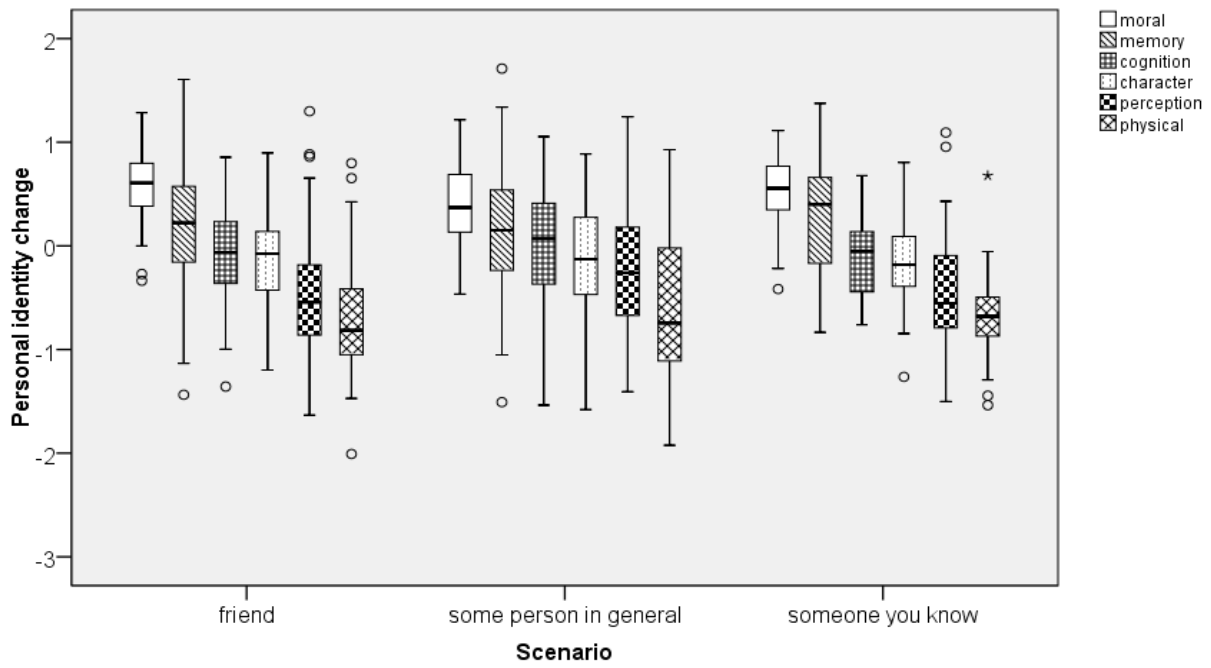
Fig. 5
The effect of scenario on the perceived change in personal identity.
The more personal the scenario, the greater the relative importance of the moral category.
The bottom end of the whisker is minimum value. The top end of the whisker is maximum value.
The bottom of the box is 25th percentile. The top of the box is 75th percentile.
The horizontal line inside the box is median. The dots represent outliers.

My Friend's True Self: Children's Concept of Personal Identity

Appendix
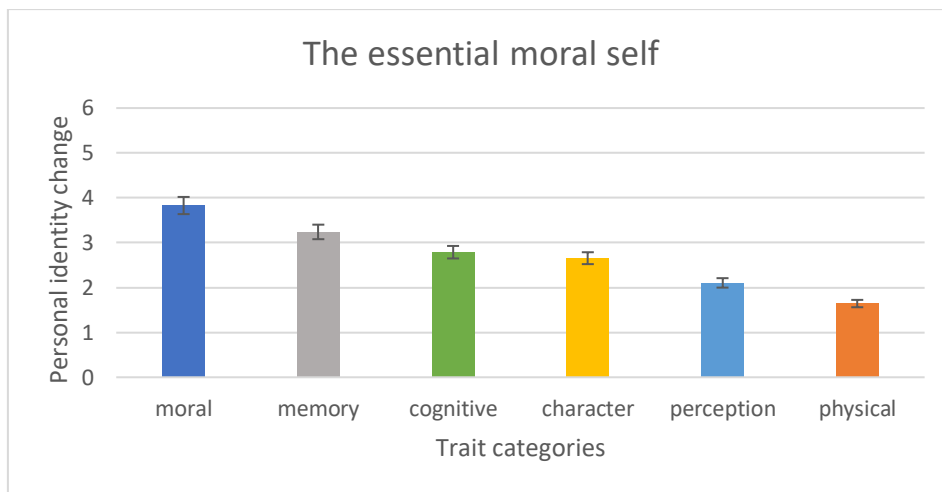
Fig. 1

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for all participants (N=217, 56,4% F) (y-axis) as they differ across six categories of change (x-axis).
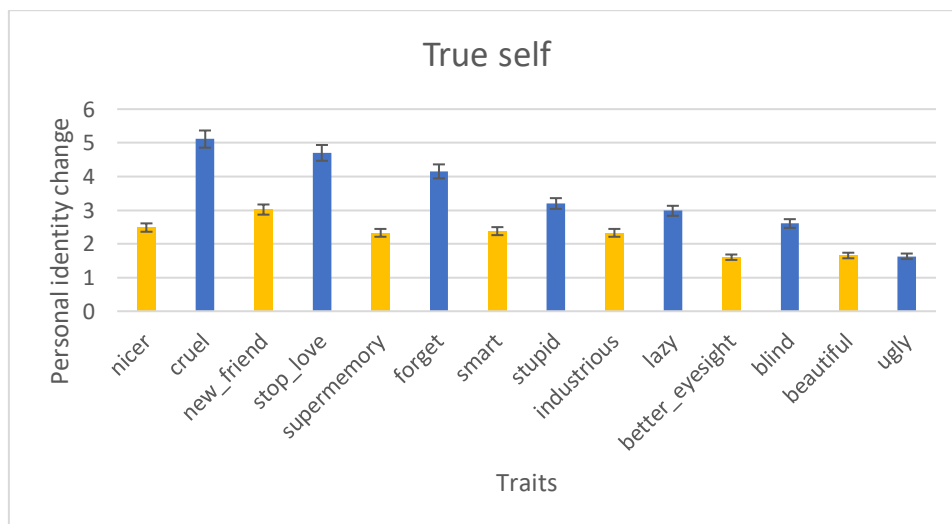
Fig. 2

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for all participants (N=217, 56,4% F) (y-axis) as they differ across all items of change and between the positive version (yellow) and the negative version (blue) of each changing trait (x-axis).
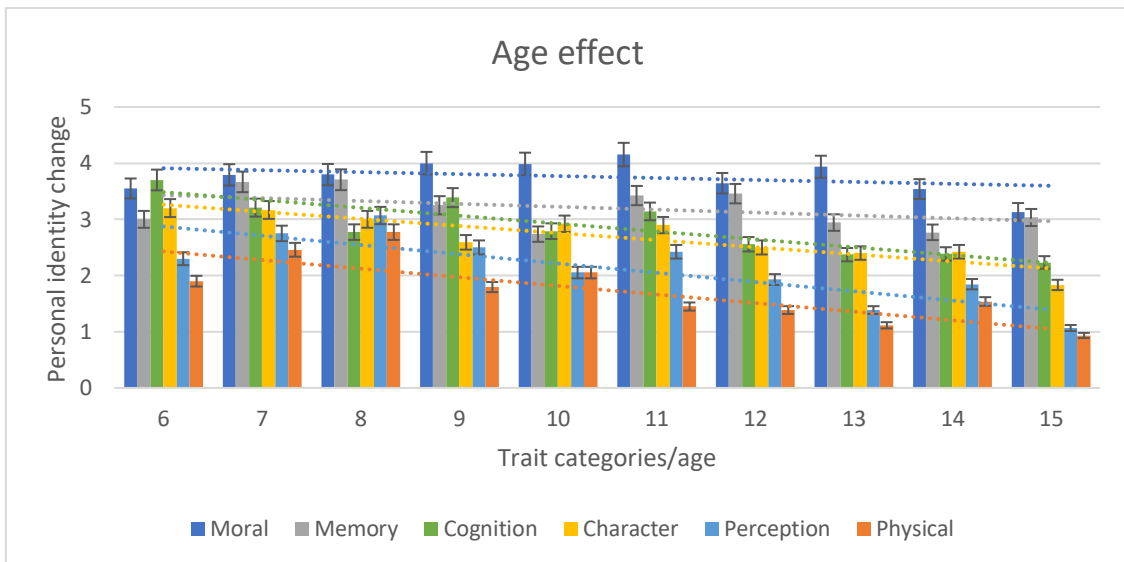
Fig. 3

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity (y-axis) as they differ across six categories of change and across different ages (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with growing age.
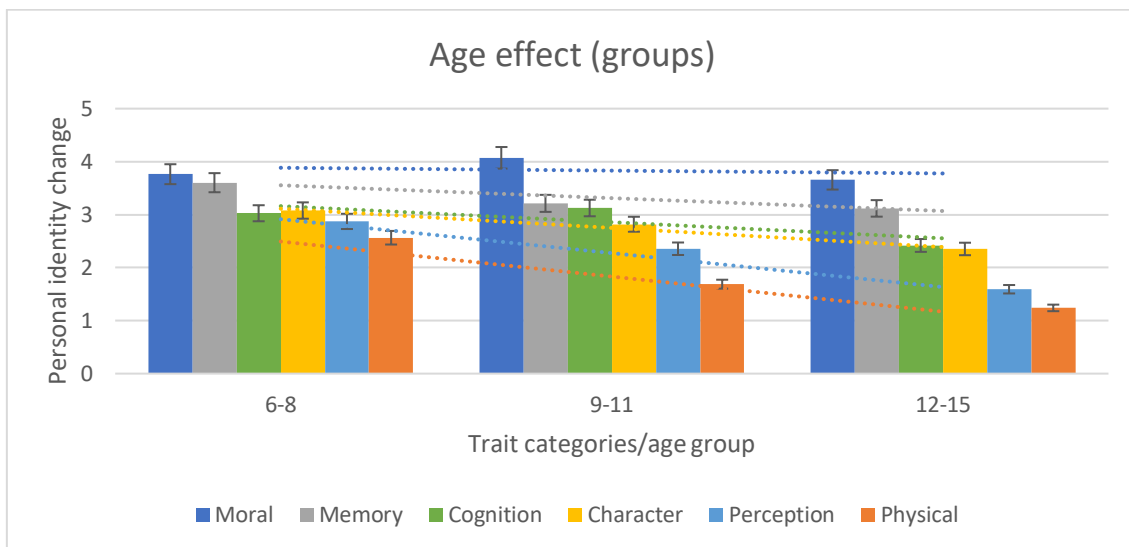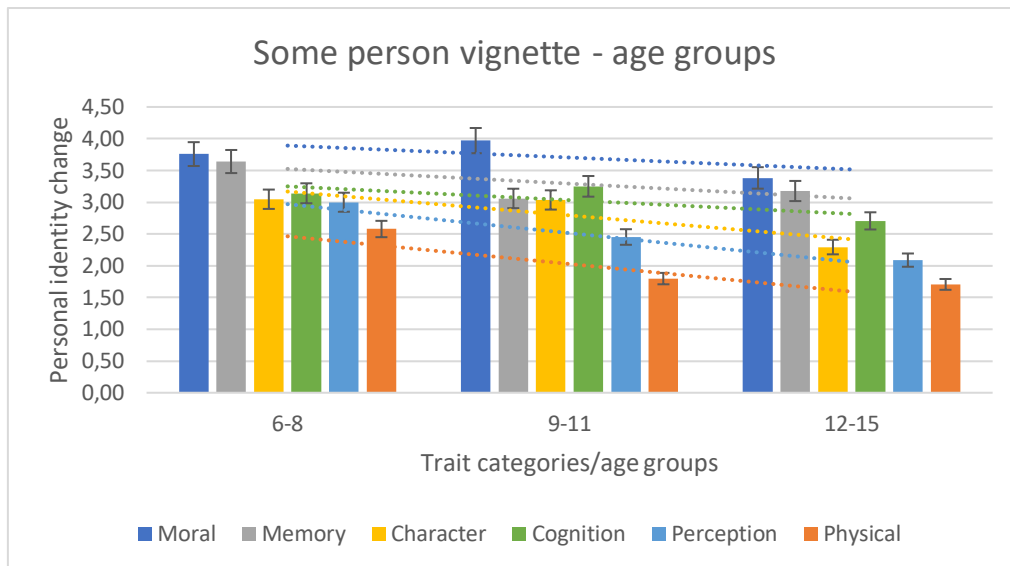

Fig. 4

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity (y-axis) as they differ across six categories of change and across different age categories (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with growing age.
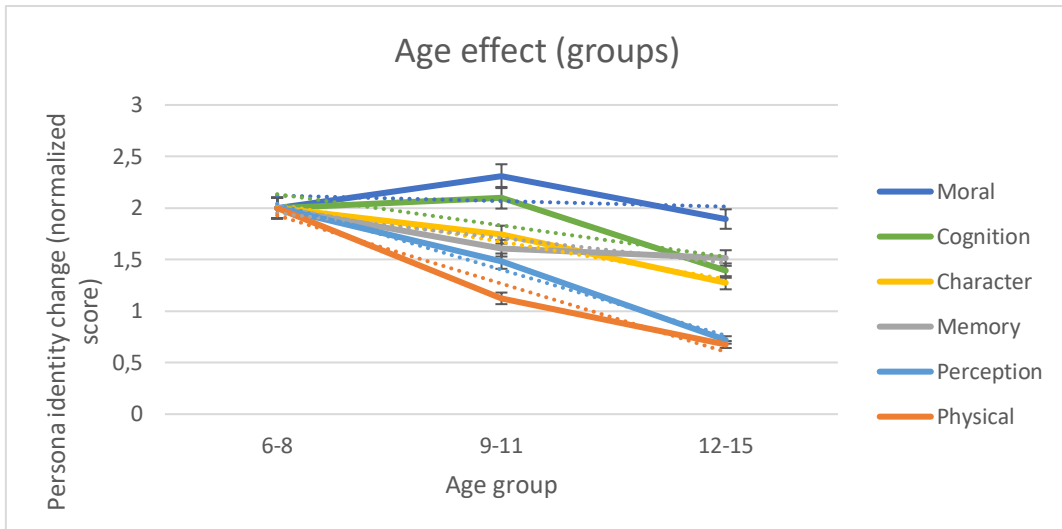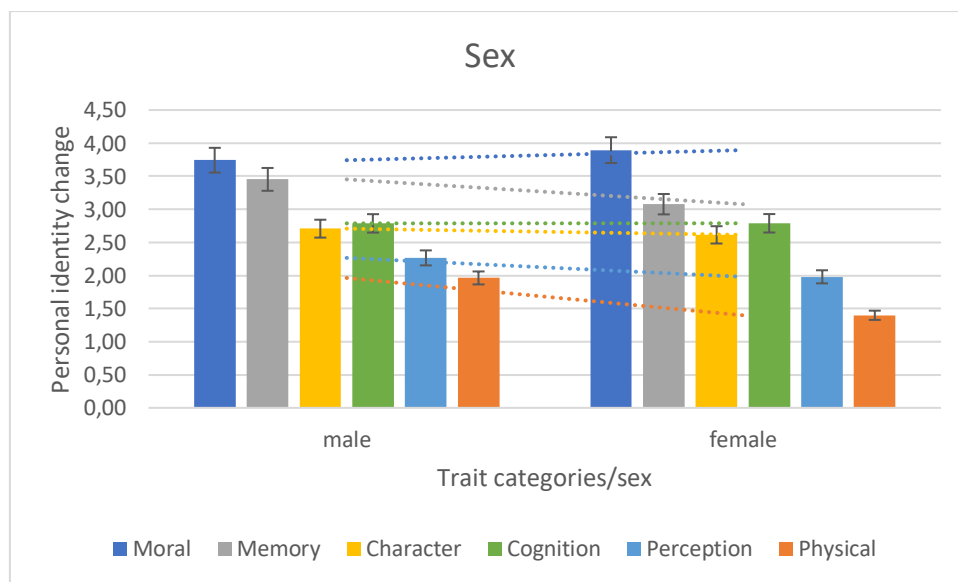
Fig. 5

Error bars: 95% CI

The graph shows normalized scores (so that they start at 2 for the lowest age) of the perceived change of personal identity (y-axis) as they differ across six categories of change and across different age categories (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with growing age.
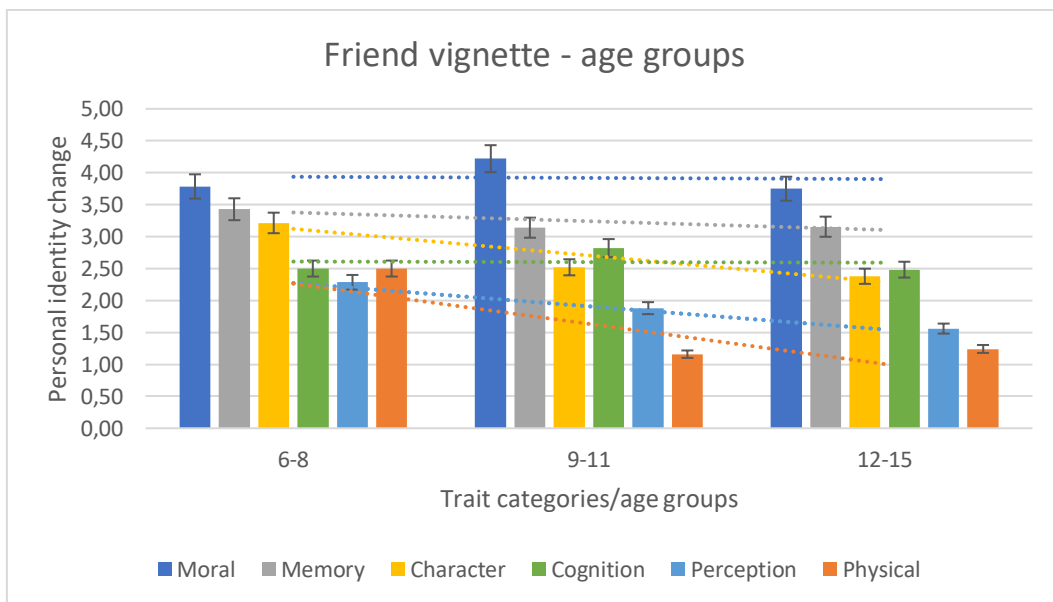
Fig. 6

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the participants answering the "friend vignette" (N=90, 64,4% F) (y-axis) as they differ across six categories of change and across different age categories (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with growing age.
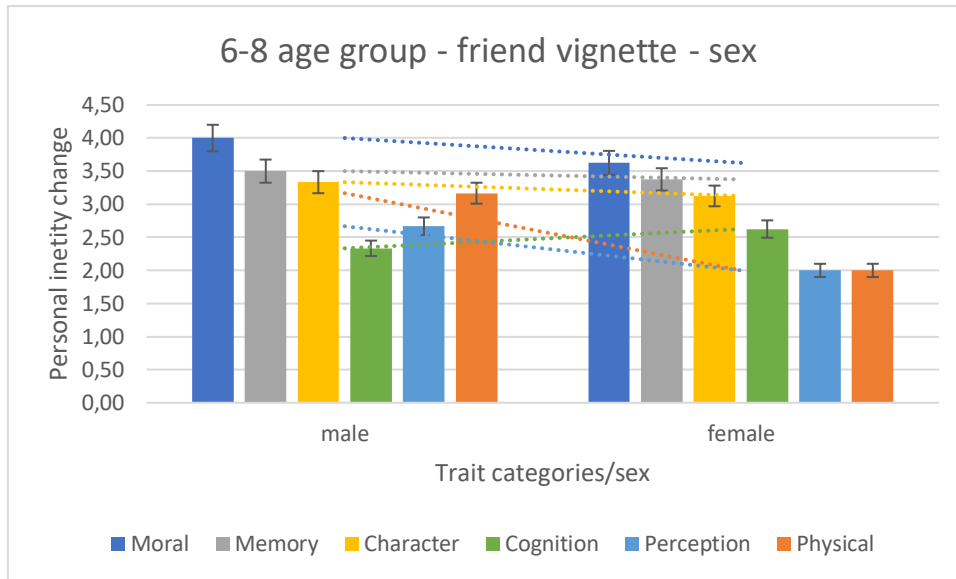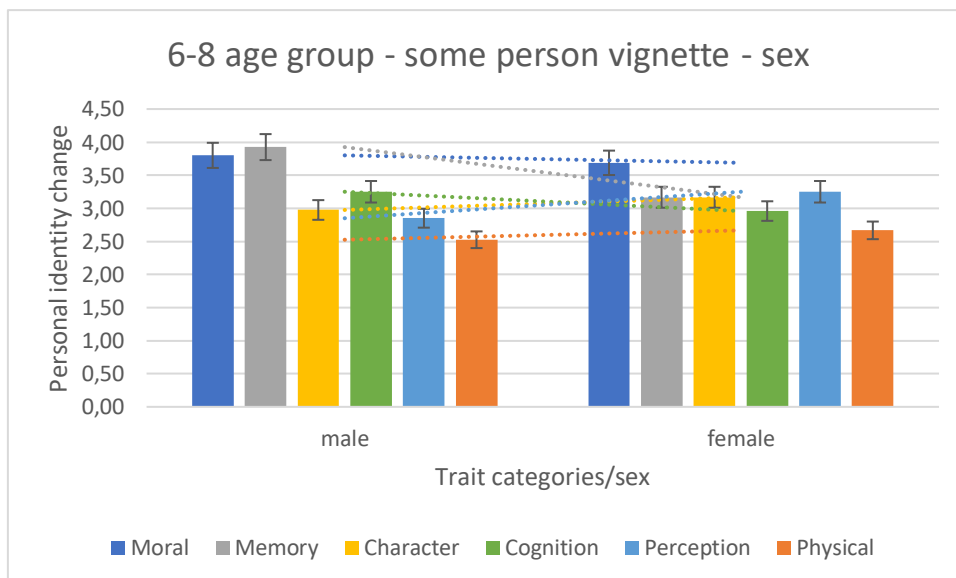
Fig. 7

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the participants answering the "some person in general" vignette (N=91, 42,9% F) (y-axis) as they differ across six categories of change and across different age categories (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with growing age.

Fig. 8

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for all participants (y-axis) as they differ across six categories of change and between sexes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score between sexes.
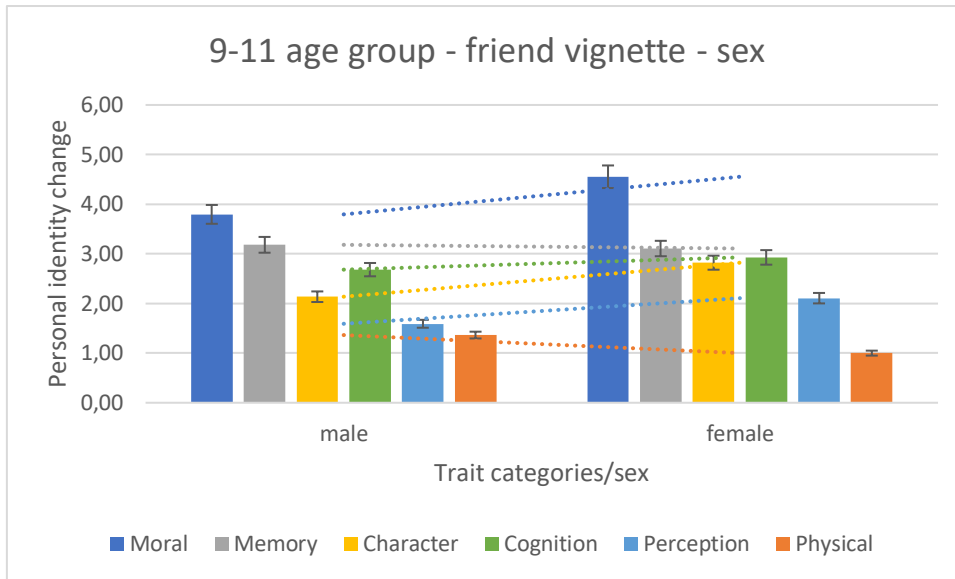
Fig. 9

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 6-8 years old participants answering the "friend" vignette (note: the sample was too small, N=7, 57,1% F) (y-axis) as they differ across six categories of change and between sexes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score between sexes.
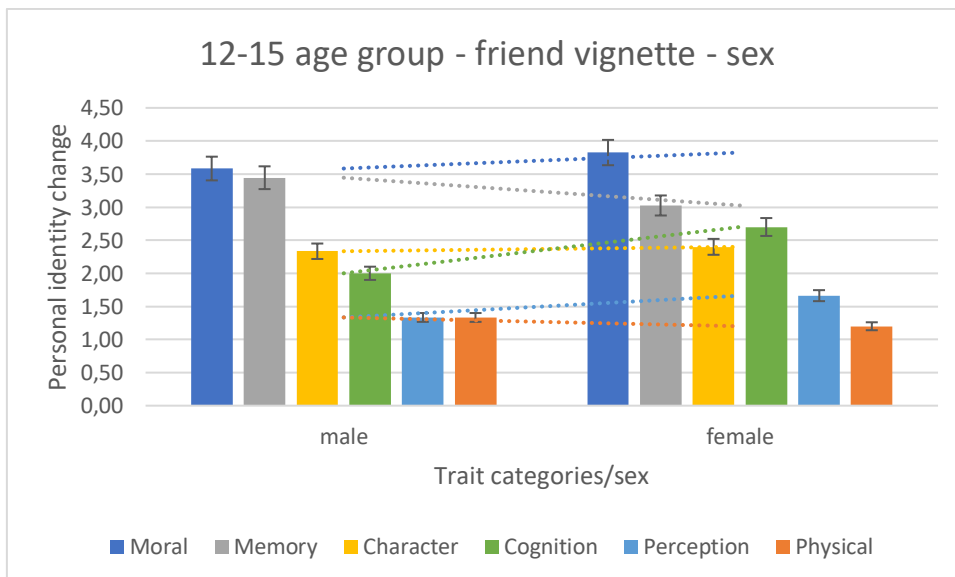

Fig. 10

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 6-8 years old participants answering the "some person in general" vignette (N=32, 37,5% F) (y-axis) as they differ across six categories of change and between sexes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score between sexes.

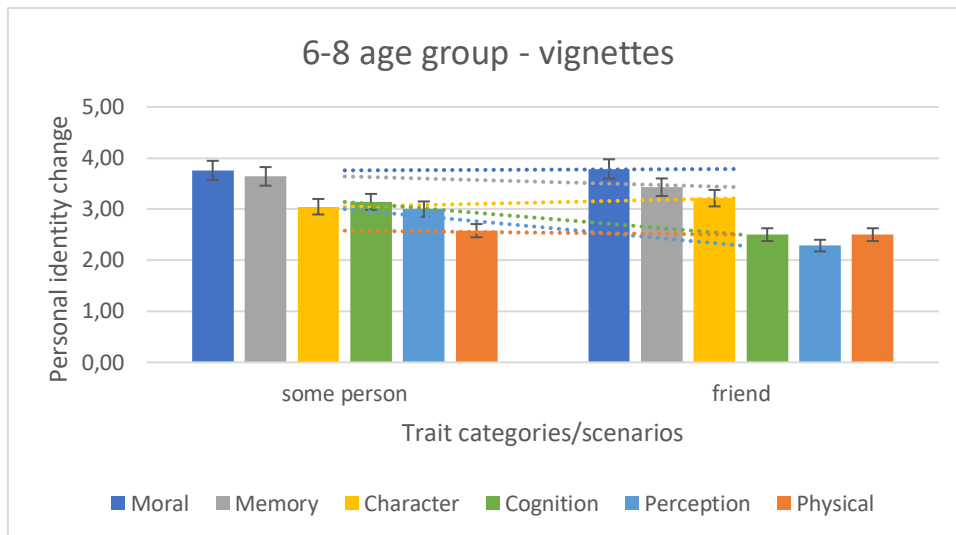**Fig. 11**

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 9-11 years old participants answering the "friend" vignette (N=25, 56% F) (y-axis) as they differ across six categories of change and between sexes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score between sexes.
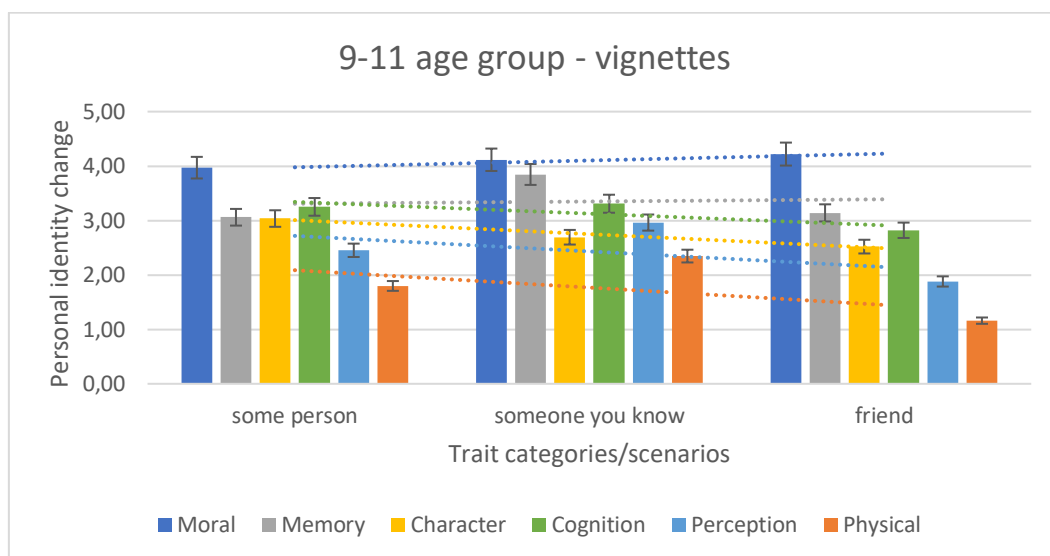

Fig. 12

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 12-15 years old participants answering the "friend" vignette (N=58, 69% F) (y-axis) as they differ across six categories of change and between sexes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score between sexes.
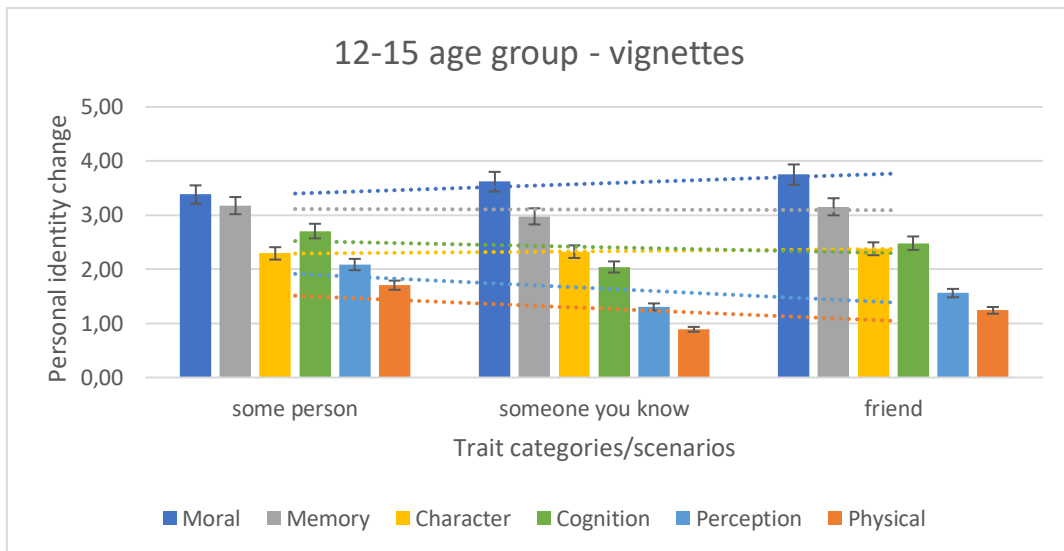
Fig. 13

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 6-8 years old participants (N=39, 41% F) (y-axis) as they differ across six categories of change and across different vignettes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with vignette change.

Fig. 14

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 9-11 years old participants (N=80, 53,8% F) (y-axis) as they differ across six categories of change and across different vignettes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with vignette change.

Fig. 15

Error bars: 95% CI

The graph shows the mean values of the original scores of the perceived change of personal identity for the 12-15 years old participants (N=98, 65,3% F) (y-axis) as they differ across six categories of change and across different vignettes (x-axis). The dotted lines show the trait categories' linear trend of the perceived personal identity change score with vignette change.

Vignette in Czech (original):

Představ si, že tvůj *kamarád /kamarád, spolužák, rovesník nebo někdo jiný /nějaký člověk* vstoupí do speciální sci-fi komory, která ho může jakkoli změnit. Jak moc podle tebe tyto možné změny zasáhnou jeho podstatu, to, co je na něm to nejdůležitější, to, kým je, to, co ho nejvíc vystihuje jako človíčka?

0 - Je to pořád on/ona, zachoval/a si to podstatné ze sebe.

6 - Už to není on/ona, ztratil/a to podstatné ze sebe.

1. Když se stane ošklivějším.
2. Když si najednou dokáže vybavit úplně všechny své vzpomínky.
3. Když přestane mít rád své kamarády a blízké.
4. Když se stane svědomitějším, pilnějším.
5. Když oslepne.
6. Když se stane krásnějším.
7. Když zapomene na všechny svoje zážitky.
8. Když se stane mnohem hodnějším k ostatním dětem.
9. Když zhloupne.
10. Když se začne přátelit s někým, koho vůbec neměl rád.
11. Když sa stane línějším.
12. Když se mu velice zlepší zrak.
13. Když se stane krutým k ostatním dětem.
14. Když zmoudří.

SCALE:

Je to pořád on/ona   Už to není on/ona

Vignette in English (translation):

Imagine that *your friend /friend, schoolmate, peer or someone else /some person* enters a special sci-fi chamber that can change them in any way. What do you think, how much will the following changes affect the person's core, that, what is most important about them, what makes them who they are, what is most characteristic of them as a person?

0 - The person is still him/her, he/she has kept the most important part of them.

6 - The person is not him/her anymore, he/she has lost the most important part of them.

1. If they become uglier.
2. If they are suddenly able to perfectly remember all their memories.
3. If they stop loving their friends and close ones.
4. If they become more industrious.
5. If they become blind.
6. If they become more beautiful (physically).
7. If they forget all their life memories.
8. If they become nicer to other kids.
9. If they become more stupid.
10. If they accept as a friend someone they didn't like before.
11. If they become lazier.
12. If their eyesight improves radically.
13. If they start to be cruel to other people.
14. If they become smarter.

SCALE:

0   **1**   **2**   **3**   **4**   **5**   **6**

It is still him/her   It is not him/her anymore